

Noise Increases Anchoring Effects



Chang-Yuan Lee¹ and Carey K. Morewedge¹

Department of Marketing, Questrom School of Business, Boston University

Psychological Science
1–16

© The Author(s) 2021

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/09567976211024254

www.psychologicalscience.org/PS



Abstract

We introduce a theoretical framework distinguishing between anchoring effects, anchoring bias, and judgmental noise: Anchoring effects require anchoring bias, but noise modulates their size. We tested this framework by manipulating stimulus magnitudes. As magnitudes increase, psychophysical noise due to scalar variability widens the perceived range of plausible values for the stimulus. This increased noise, in turn, increases the influence of anchoring bias on judgments. In 11 preregistered experiments ($N = 3,552$ adults), anchoring effects increased with stimulus magnitude for point estimates of familiar and novel stimuli (e.g., reservation prices for hotels and donuts, counts in dot arrays). Comparisons of relevant and irrelevant anchors showed that noise itself did not produce anchoring effects. Noise amplified anchoring bias. Our findings identify a stimulus feature predicting the size and replicability of anchoring effects—stimulus magnitude. More broadly, we show how to use psychophysical noise to test relationships between bias and noise in judgment under uncertainty.

Keywords

anchoring bias, noise, judgment under uncertainty, numerical cognition, open data, open materials, preregistered

Received 7/7/20; Revision accepted 5/5/21

An anchoring effect occurs when people consider one number (an anchor) and their subsequent judgments are assimilated to it. Anchoring effects occur in judgments ranging from mundane trivia answers to the selling price of homes, but they are not ubiquitous (Northcraft & Neale, 1987; Tversky & Kahneman, 1974). Anchoring effects vary in size across anchors and judges and contexts (Jacowitz & Kahneman, 1995; Jung et al., 2016; Smith et al., 2013; Wilson et al., 1996). We propose that noise in the mental representation of the stimulus estimated (i.e., the target) is a critical determinant of the size of anchoring effects.

We base our theoretical framework on the anchoring-and-adjustment heuristic (Epley & Gilovich, 2006; Simmons et al., 2010; Tversky & Kahneman, 1974). It posits that people estimate the value of a target (e.g., the duration of Mars's orbit) by identifying an anchor (e.g., Earth's orbit = 365 days). People then adjust from that anchor until they reach a range of plausible values for the point estimate and stop at a value within that range. Because adjustment is usually insufficient (i.e., people typically stop at a value before rather than beyond the correct answer), estimates are biased by consideration of anchors (Quattrone et al., 1984). The

average estimate of Mars's orbit is 492 days, for instance, which is 195 days fewer than the right answer (i.e., 687 days; Epley & Gilovich, 2006).

The terms *anchoring effect* and *anchoring bias* are used interchangeably to describe this biasing effect of anchors (e.g., Chapman & Johnson, 1994; Englich et al., 2006; Epley & Gilovich, 2006), but errors in judgment are driven by both bias and noise (Kahneman et al., 2016). We propose that the terms should be distinct because anchoring effects are also a product of bias and noise. Most anchoring research examines an anchoring effect, which we define as the absolute effect of an anchor on an estimate. It can be calculated as the raw difference between point estimates influenced by low and high anchors or as the raw difference between point estimates made with and without an anchor. In Figure 1, the anchoring effect indicated by the red line in example A is the raw difference between point estimates made with low and high anchors, A_L and A_H .

Corresponding Author:

Carey K. Morewedge, Boston University, Questrom School of Business, Department of Marketing
Email: morewedg@bu.edu

Noise in our framework reflects the width of the range of plausible values for the point estimate, that is, the distance between the minimum and maximum plausible values. It is essentially the judge's subjective confidence interval (CI). In Figure 1, the red line in example C indicates the plausible range width for point estimates C_L and C_H . *Anchoring bias* is the degree of undercorrection from the anchor in the point estimate relative to the range of plausible values. In example B in Figure 1, judges corrected only halfway from plausible extremums to the midpoint in point estimates B_L and B_H . Anchoring bias can be measured with a skew index (Epley & Gilovich, 2006) dividing (a) the difference between a point estimate and the plausible extremum nearest to the anchor (low anchor \rightarrow minimum plausible value; high anchor \rightarrow maximum plausible value) by (b) the difference between the minimum and maximum plausible value (for details, see Experiment 6a). Thus, the same anchoring bias produces a smaller anchoring effect when there is less noise and a larger anchoring effect when there is more noise.

Our framework shows how noise and anchoring bias together determine the size of anchoring effects. Furthermore, it helps specify whether anchoring effects vary across factors such as judges, anchors, and contexts because they modulate anchoring bias or noise. We illustrate these proposals in Figure 1. An expert (example A) and novice (example C) could be equally

Statement of Relevance

How much is a house worth? Homebuyers and even experienced realtors are biased by the list price when estimating home values: The higher the list price, the greater the presumed worth. This happens because people insufficiently adjust from the first value considered (the list price) in the subsequent judgment (the value of the home). Their estimates are biased by the "anchor" of the list price. The size of anchoring effects is influenced by the amount of adjustment. People usually stop adjusting their estimates too early, at the first reasonable value. We tested a complementary influence: noise (e.g., the width of the range of values that seem reasonable). The effect of noise can be seen in the major finding of this research: that anchoring effects are larger when people estimate larger numbers (which have a larger range) than smaller numbers (which have a smaller range). Importantly, anchoring effects do not apply just to house prices. Anchoring effects influence all kinds of everyday and consequential judgments.

uncertain about the plausible value of a stimulus (same noise), but the expert exhibits a smaller anchoring effect than the novice because they are less biased by

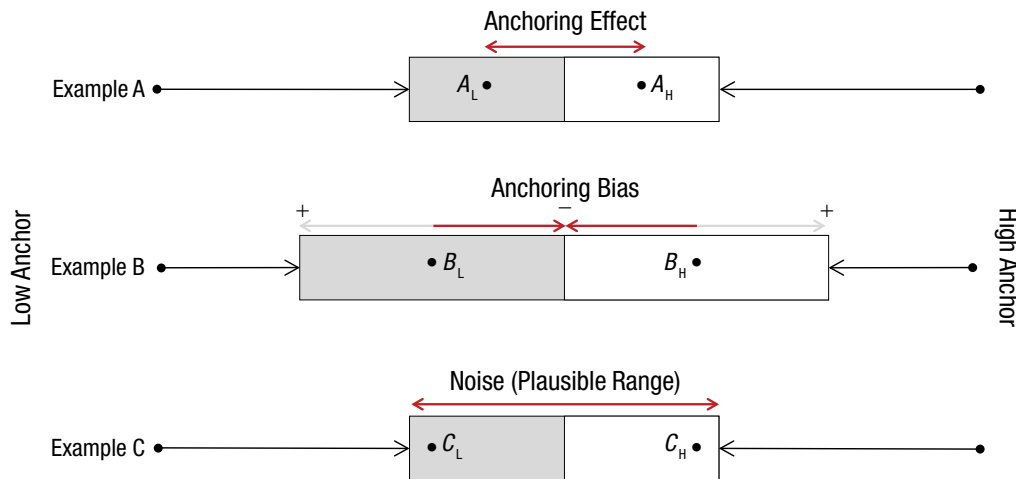


Fig. 1. Anchoring effects, anchoring bias, and judgmental noise. The anchoring effect (indicated by the red line in example A) is the raw difference between point estimates made with low and high anchors, A_L and A_H . Anchoring bias is the degree of undercorrection from the anchor in the point estimate relative to the range of plausible values. In example B, judges corrected only halfway from plausible extremums to the midpoint in point estimates B_L and B_H . Noise in our framework reflects the width of the range of plausible values for the point estimate, that is, the distance between the minimum and maximum plausible values. The red line in example C indicates the plausible range width for point estimates C_L and C_H . The (absolute) anchoring effects of low and high anchors on point estimates are smallest in example A ($A_H - A_L$) and equally larger in examples B and C ($B_H - B_L$; $C_H - C_L$). Anchoring bias is greatest in example C, where there is the least (relative) correction from anchors, and equally smaller in examples A and B. Noise is greatest in example B, which has the widest plausible range of stimulus values, and equally smaller in examples A and C. Black arrows depict adjustment from low and high anchors to the range of plausible values of the point estimate. Red arrows depict the range of the effect in question.

the anchor (less anchoring bias). Alternatively, the expert (example A) could be as biased by the anchor as the novice (example B; same anchoring bias), but they exhibit a smaller anchoring effect because the range of plausible values that they consider is narrower than the range the novice considers (less noise).

We tested our theory by manipulating anchors and stimulus magnitudes. Scalar variability makes mental representations of numbers noisier as stimulus magnitudes increase (Feigenson et al., 2004). The range of plausible values for point estimates should then widen with stimulus magnitude. The range between the minimum and maximum plausible weight of a small dog, for instance, should be narrower than the range between the minimum and maximum plausible weight of a large dog. We tested this prediction in six pretests. Consequently, anchoring effects should increase with stimulus magnitude, even when anchoring bias is similar in estimates of smaller and larger targets (we tested this in Experiments 6a and 6b).

We tested whether anchoring effects increase with stimulus magnitude in Experiments 1a to 3b. In Experiments 4a and 4b, we tested whether low stimulus magnitudes explain cases in which anchoring effects are weak or have failed to replicate (Jung et al., 2016; Maniadis et al., 2014). In Experiment 5, we tested the proposed relationship between anchoring effects, anchoring bias, and noise. In Experiments 6a and 6b, we directly compared the effects of stimulus magnitude on anchoring effects, anchoring bias, and noise.

All experiments were preregistered on AsPredicted (for links to the preregistrations, see the Open Practices section). We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in all experiments. Generally, we aimed to recruit at least 100 participants per condition in our experiments because the focal statistical predictions were interaction effects. The Boston University Institutional Review Board for the Charles River Campus (Protocol No. 3626E) approved the use of human subjects in all experiments. All data and materials are available at <https://osf.io/9xun6/>.

Pretests: Effect of Stimulus Magnitudes on Plausible Range Widths

Method

In six categories of stimuli, we tested whether the plausible range of stimulus values widens as stimulus magnitudes increase.

Miami hotels. We recruited 50 participants from Amazon Mechanical Turk (MTurk) so the study would be

well powered to detect medium-size effects within subjects, and all 50 completed the pretest. Participants saw pictures of three hotels in Miami Beach, Florida, vertically differentiated by star rating (i.e., a two-star hotel, a three-star hotel, and a four-star hotel; see Fig. S1 in the Supplemental Material). In open-ended response boxes, participants estimated the highest and lowest market price (U.S. dollars per night) of a standard room in each hotel in the past year.

Dog breeds. We recruited 50 participants from MTurk, and 48 completed the pretest. Participants saw pictures of three adult dogs of different breeds, which they were told varied in size from small to medium to large (i.e., basenji, American Staffordshire terrier, and Bernese mountain dog, respectively; see Fig. S2 in the Supplemental Material). In open-ended response boxes, participants estimated the maximum and minimum plausible weight in pounds of each of the three dog breeds.

French fries. We recruited 50 participants from MTurk, and 52 completed the pretest. Participants saw pictures of small, medium, and large servings of McDonald's french fries (see Fig. S3 in the Supplemental Material). In open-ended response boxes, participants estimated the maximum and minimum plausible number of calories in each serving.

Dot arrays. We recruited 50 participants from MTurk, and all completed the pretest. Participants saw pictures of three dot arrays that obviously varied in the number of dots that each contained (i.e., 35, 97, and 273 dots; see Fig. S4 in the Supplemental Material). In open-ended response boxes, participants estimated the maximum and minimum plausible number of dots in each of the three arrays.

Donuts. We recruited 50 participants from MTurk, and all completed the pretest. Participants were first presented with an image of Dream Fluff Donuts, adapted from the study by Jung and colleagues (2016). In open-ended response boxes, participants estimated the highest and lowest market price for one donut and for one dozen donuts made by Dream Fluff Donuts.

Unpleasant tones. We recruited 50 participants from MTurk, and all completed the pretest. Participants first listened to an unpleasant tone for 30 s, the same tone used by Maniadis and colleagues (2014). Participants then read that another 100 MTurk workers had reported the maximum amount of money they were willing to accept to listen to the same tone for 60 s, 180 s, and 300 s. In open-ended response boxes, participants estimated

Table 1. Mean Plausible Range Width for Each Stimulus Magnitude in the Six Pretests

| Stimuli | Stimulus magnitude | | |
|--|---------------------------|----------------------------|----------------------------|
| | Small | Medium | Large |
| Miami hotels (U.S. dollars per night) | $\Delta 89.26_a$ (102.01) | $\Delta 115.94_b$ (90.56) | $\Delta 144.22_c$ (115.33) |
| Dog breeds (pounds) | $\Delta 29.40_a$ (18.48) | $\Delta 53.77_b$ (48.94) | $\Delta 74.75_c$ (51.88) |
| French fries servings (calories) | $\Delta 81.98_a$ (119.79) | $\Delta 117.00_b$ (178.96) | $\Delta 156.38_c$ (254.52) |
| Dot arrays (counts) | $\Delta 21.64_a$ (13.29) | $\Delta 40.06_b$ (29.97) | $\Delta 118.20_c$ (148.95) |
| Donuts (U.S. dollars) | $\Delta 1.75_a$ (1.94) | | $\Delta 5.77_b$ (5.39) |
| Unpleasant tones (U.S. dollars) | $\Delta 3.24_a$ (5.08) | $\Delta 4.73_b$ (6.51) | $\Delta 7.38_c$ (9.37) |

Note: Means within rows that do not share subscripts differ significantly ($p < .05$), as determined by a paired-samples t test (within subjects). Standard deviations are given in parentheses.

the highest and lowest amounts of money that those MTurk workers requested to listen to the tone for each of the three durations.

Results

We first computed plausible range widths for each stimulus, within each participant, by subtracting the minimum estimate from the maximum estimate; mean plausible range widths and standard deviations are reported in Table 1. We then compared the widths of plausible ranges for stimuli of the largest, medium, and smallest magnitude. As predicted, the mean plausible range of the stimulus with the largest magnitude was wider than the plausible range of the stimuli with the medium and smallest magnitudes (all $t_s \geq 2.16$, all $p_s \leq .036$, all $d_s \geq 0.31$), and the plausible range of the stimulus with the medium magnitude was wider than that of the stimulus with the smallest magnitude (all $t_s \geq 2.75$, all $p_s \leq .008$, all $d_s \geq 0.39$). For exact values for all comparisons, see Section S2 in the Supplemental Material.

Experiments 1a to 3b: Directional Tests

Given that plausible ranges of stimulus values widen with stimulus magnitude, our theory predicts that anchoring effects should increase with stimulus magnitude. To compare across subjective and objective judgments, we operationalized anchoring effects as the difference in point estimates between participants exposed to a low anchor, high anchor, or no anchor (depending on the experiment). In Experiments 1a and 1b, we manipulated externally provided anchors between subjects and targets within subjects. In Experiments 2a and 2b, we manipulated internally generated anchors between subjects and targets within subjects.

In Experiments 3a and 3b, we manipulated externally provided anchors and targets between subjects.

Experiment 1a: willingness to pay for Miami hotels

Method.

Participants and design. We requested 300 participants from MTurk, and 297 completed the experiment (39% female; age: $M = 36.40$ years, $SD = 10.43$). In a mixed, between-subjects design, we randomly assigned each participant to one of three externally-provided-anchor conditions: control (i.e., no anchor), low anchor, or high anchor. Each participant then reported the maximum amount they were willing to pay for three Miami hotels (within subjects).

Procedure. In the no-anchor condition, participants saw no anchor. They imagined purchasing a hotel room for one night during an upcoming trip to Miami and were presented with the name, a photograph, a TripAdvisor traveler rating, and a star rating for each of three Miami Beach hotels (i.e., a two-star hotel, a three-star hotel, and a four-star hotel). Using an open-ended response box, participants then reported the maximum amount they would be willing to pay in U.S. dollars for a room for one night at each of the three hotels. Values for all three hotels were elicited simultaneously on one survey page.

In the low-anchor condition, participants were first informed of the price of a room for one night in a one-star Miami Beach hotel (i.e., priced at \$44). In the high-anchor condition, participants were first informed of the price of a room for one night in a five-star Miami Beach hotel (i.e., priced at \$610). Participants then reported the maximum amount they would be willing to pay for a room (per night) for each of the three hotels, just as controls.

Results. To test our directional predictions, we first examined how much participants were willing to pay for the three target hotels in a 3 (anchor: no, low, high; between subjects) \times 3 (hotel: two star, three star, four star; within subjects) mixed analysis of variance (ANOVA), which revealed a significant main effect of anchor, $F(2, 294) = 37.37, p < .001, \eta_p^2 = .20$, and a significant main effect of hotel, $F(1, 294) = 317.44, p < .001, \eta_p^2 = .52$. More important, these main effects were qualified by a significant Anchor \times Hotel interaction, $F(2, 294) = 34.18, p < .001, \eta_p^2 = .19$.¹ Means and standard deviations are reported in Figure 2 (see also Table S1 in the Supplemental Material).

We decomposed the interaction with comparisons across each pair of conditions in separate mixed ANOVAs. Most important, comparing the low- and high-anchor conditions revealed a significant 2 (anchor: low, high) \times 3 (hotel: two star, three star, four star) interaction, $F(1, 192) = 51.23, p < .001, \eta_p^2 = .21$. Simple comparisons showed that participants were willing to pay more for the four-star hotel in the high-anchor condition than in the low-anchor condition, $t(192) = 9.29, 95\% \text{ CI for the mean difference} = [144.53, 222.44], p < .001, d = 1.33$. Participants were also willing to pay more for the two-star hotel in the high-anchor condition than in the low-anchor condition, $t(192) = 4.55, 95\% \text{ CI for the mean difference} = [35.75, 90.51], p < .001, d = 0.65$. Moreover, a 2 (anchor: low, high) \times 2 (hotel: two star, four star) interaction revealed that this difference between conditions was significantly greater for the four-star than the two-star hotel, $F(1, 192) = 59.08, p < .001, \eta_p^2 = .24$ (Nieuwenhuis et al., 2011).

Comparing the no-anchor and high-anchor conditions revealed a significant 2 (anchor: no, high) \times 3 (hotel: two star, three star, four star) interaction, $F(1, 199) = 31.72, p < .001, \eta_p^2 = .14$. Simple comparisons showed that participants were willing to pay more for the four-star hotel in the high-anchor condition than in the no-anchor condition, $t(199) = 6.61, 95\% \text{ CI for the mean difference} = [93.19, 172.45], p < .001, d = 0.93$. Participants were also willing to pay more for the two-star hotel in the high-anchor condition than in the no-anchor condition, $t(199) = 2.33, 95\% \text{ CI for the mean difference} = [5.09, 61.75], p = .021, d = 0.33$. Moreover, a 2 (anchor: no, high) \times 2 (hotel: two star, four star) interaction revealed that the difference between conditions was significantly greater for the four-star than the two-star hotel, $F(1, 199) = 38.26, p < .001, \eta_p^2 = .16$.

Comparing the no-anchor and low-anchor conditions revealed a marginal 2 (anchor: no, low) \times 3 (hotel: two star, three star, four star) interaction, $F(1, 197) = 3.48, p = .06, \eta_p^2 = .02$. Simple comparisons showed that participants were willing to pay more for the four-star hotel in the no-anchor condition than in the low-anchor condition, $t(197) = 3.90, 95\% \text{ CI for the mean difference} =$

$[25.04, 76.30], p < .001, d = 0.55$. Participants were also willing to pay more for the two-star hotel in the no-anchor condition than in the low-anchor condition, $t(197) = 2.40, 95\% \text{ CI for the mean difference} = [5.29, 54.13], p = .017, d = 0.34$. Moreover, a 2 (anchor: no, low) \times 2 (hotel: two star, four star) interaction revealed that this difference between conditions was significantly greater for the four-star than the two-star hotel, $F(1, 197) = 4.33, p = .039, \eta_p^2 = .02$.

Experiment 1b: weight of dog breeds

Method.

Participants and design. We requested 200 participants from MTurk, and 201 completed the experiment (43% female; age: $M = 37.03$ years, $SD = 11.49$). In a mixed design, we randomly assigned each participant to a low- or high-externally-provided-anchor condition (between subjects). Each participant then made weight estimates for three dog breeds (within subjects).

Procedure. Participants randomly assigned (between subjects) to the low-anchor condition first saw a picture of an adult Australian terrier and were told that the average weight of its breed is 12 lb. In the high-anchor condition, participants first saw a picture of an adult Boerboel and read that the average weight of its breed is 200 lb. In open-ended response boxes appearing on the same page, participants then estimated the average weight, in pounds, of an adult basenji, an American Staffordshire terrier, and a Bernese mountain dog. A picture of each dog accompanied the breed name. The ranking of the weights of the dog breeds was made explicit. In the low-anchor condition, participants saw the following ranking (from lowest to highest ranked): Australian terrier < basenji < American Staffordshire terrier < Bernese mountain dog. In the high-anchor condition, they saw the following ranking (from lowest to highest ranked): basenji < American Staffordshire terrier < Bernese mountain dog < Boerboel.

Results. Thirty-five participants were excluded from all analyses because they gave weight estimates that were inconsistent with the explicit ranking of the weights of the dog breeds (a preregistered exclusion criterion). To test our directional predictions, we examined weight estimates of the three target dog breeds in a 2 (anchor: low, high; between subjects) \times 3 (breed: basenji, American Staffordshire terrier, Bernese mountain dog; within subjects) mixed ANOVA, which revealed a significant main effect of anchor, $F(1, 164) = 70.61, p < .001, \eta_p^2 = .30$, and a significant main effect of breed, $F(1, 164) = 836.88, p < .001, \eta_p^2 = .84$. The main effects were qualified by a significant Anchor \times Breed interaction, $F(1, 164) = 45.27, p < .001, \eta_p^2 = .22$, which revealed that the anchoring effect

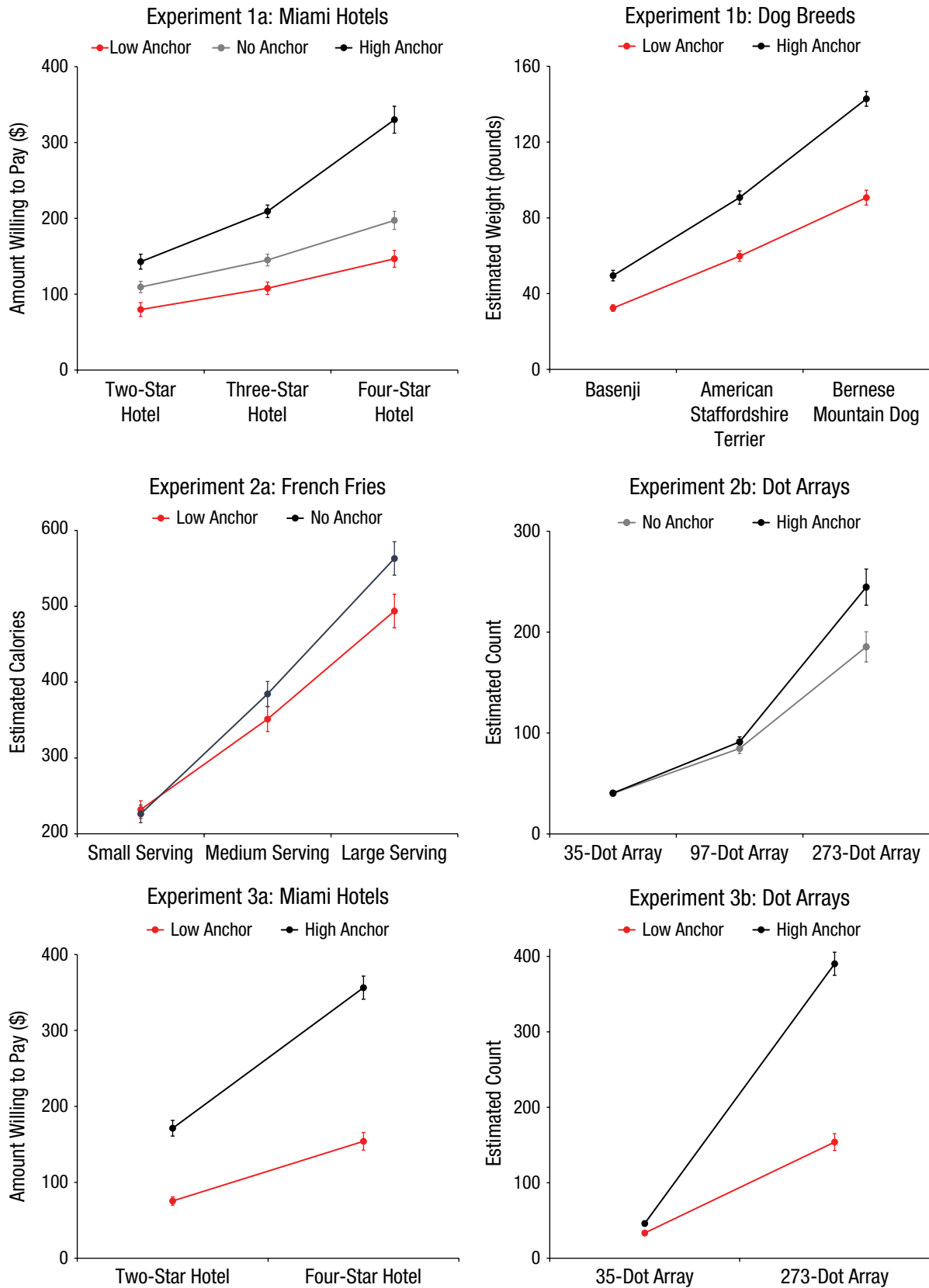


Fig. 2. Mean point estimate (y-axis) for each stimulus magnitude (x-axis) as a function of anchor condition in Experiments 1a to 3b. In separate experiments, participants indicated how much they would be willing to pay for hotels (Experiments 1a and 3a), how much they estimated dog breeds weigh (Experiment 1b), how many calories they estimated were in servings of french fries (Experiment 2a), and how many dots they estimated were in an array (Experiments 2b and 3b). Error bars represent ± 1 SEM.

increased with stimulus magnitude. Weight estimates for each breed were heavier in the high- than in the low-anchor conditions, basenji: $t(164) = 5.29, p < .001$, 95% CI for the mean difference = [10.75, 23.52], $d = 0.82$; American Staffordshire terrier: $t(164) = 6.93, p < .001$, 95% CI for the mean difference = [22.12, 39.75], $d = 1.07$; Bernese mountain dog: $t(164) = 9.40, p < .001$, 95% CI for the mean difference = [41.23, 63.15], $d = 1.46$. A significant 2 (anchor: high, low) \times 2 (breed: basenji, Bernese mountain dog) interaction revealed that this difference in weight estimates between anchor conditions was greater for Bernese mountain dogs than basenjis, $F(1, 164) = 57.39, p < .001, \eta_p^2 = .26$. Means and standard deviations are reported in Figure 2 (see also Table S1).

Experiment 2a: calories in servings of McDonald's french fries

Method.

Participants and design. We requested 200 participants from MTurk, and 211 completed the experiment (48% female; age: $M = 37.68$ years, $SD = 11.65$). In a mixed design, we randomly assigned each participant to a low- or no-internally-generated-anchor condition (between subjects). Each participant then made calorie estimates for three servings of McDonald's french fries (within subjects).

Procedure. At the beginning of the experiment, all participants were told that McDonald's offered four different servings of french fries: kids, small, medium, and large. In the no-anchor condition, participants saw pictures of three different servings of McDonald's french fries (i.e., small, medium, and large) and estimated the number of calories in each serving on the same page. On a separate page, they next saw a picture of a serving of McDonald's kids french fries and estimated the number of calories in that serving. Participants in the low-anchor condition first saw and estimated the calories contained in the serving of kids french fries. They then saw and estimated the calories contained in small, medium, and large servings. Calorie estimates were made in open-ended response boxes.

Finally, participants indicated whether they searched online for the calorie information when estimating the numbers of calories in McDonald's french fries.

Results. Fourteen participants who reported searching online for the calorie information were excluded from the analysis (this exclusion criterion was preregistered).

Anchor. Calorie estimates for the serving of kids french fries were significantly higher in the low-anchor condition ($M = 184.04, SD = 108.24$) than in the no-anchor condition ($M = 150.11, SD = 78.11$), $t(195) = 2.52$, 95% CI for the mean difference = [7.38, 60.47], $p = .01, d = 0.36$.

Targets. To test our directional predictions, we examined calorie estimates for the three adult-size servings in a 2 (anchor: no, low; between subjects) \times 3 (serving size: small, medium, large; within subjects) mixed analysis of covariance (ANCOVA) with calorie estimate for kids french fries as a covariate. It revealed a significant main effect of anchor, $F(1, 194) = 24.36, p < .001, \eta_p^2 = .11$, and a main effect of serving size, $F(1, 194) = 76.60, p < .001, \eta_p^2 = .28$. More important, there was a significant Anchor \times Serving Size interaction, $F(1, 194) = 23.90, p < .001, \eta_p^2 = .11$, which held when calorie estimates for kids french fries were not included as a covariate, $F(1, 195) = 11.36, p = .001, \eta_p^2 = .06$.

Simple comparisons showed that calorie estimates for the large serving were significantly lower in the low-anchor condition than in the no-anchor condition, $t(195) = 2.03$, 95% CI for the mean difference = [1.99, 136.76], $p = .04, d = 0.29$. By contrast, there was no significant difference in calorie estimates for the small or medium servings between the low-anchor condition and the no-anchor condition, small serving: $t(195) = 0.33, p = .74$, 95% CI for the mean difference = [-39.94, 28.51], $d = 0.04$; medium serving: $t(195) = 1.30, p = .19$, 95% CI for the mean difference = [-17.06, 83.37], $d = 0.19$. A significant 2 (anchor: no, low; between subjects) \times 2 (serving size: small, large; within subjects) interaction revealed that the anchoring effect was significantly greater for the large than for the small serving of french fries, $F(1, 195) = 12.05, p = .001, \eta_p^2 = .06$. Means and standard deviations are reported in Figure 2 (see also Table S1).

Experiment 2b: counts in dot arrays

Method.

Participants and design. We requested 200 participants from MTurk, and 200 completed the experiment (46% female; age: $M = 41.01$ years, $SD = 13.37$). In a mixed design, we randomly assigned each participant to a no- or high-internally-generated-anchor condition (between subjects). Each participant then made dot estimates for three related stimuli (within subjects).

Procedure. In the no-anchor condition, participants saw no anchor. They estimated the number of dots in 35-, 97-, and 273-dot arrays using three open-ended response boxes that appeared on the same page. The number of dots in each of these three arrays was not disclosed to participants. In the high-anchor condition, participants first estimated the number of dots in a 500-dot array in an open-ended response box. On a separate page, each participant next estimated the number of dots in 35-, 97-, and 273-dot arrays using three open-ended response boxes that appeared on the same page. All dots were the same size.

Results.

Anchor. The mean dot estimate of the anchor dot array was 297.72 ($SD = 338.90$).

Targets. To test our directional predictions, we examined dot estimates for the three target dot arrays in a 2 (anchor: no, high; between subjects) \times 3 (dot array: 35, 97, 273; within subjects) mixed ANOVA, which revealed significant main effects of anchor, $F(1, 198) = 4.43$, $p = .036$, $\eta_p^2 = .02$, and of dot array, $F(1, 198) = 235.79$, $p < .001$, $\eta_p^2 = .54$. More important, there was a significant Anchor \times Dot Array interaction, $F(1, 198) = 7.59$, $p < .01$, $\eta_p^2 = .04$, which suggests that the anchoring effect increased with stimulus magnitude. Simple comparisons showed that dot estimates for the 273-dot array were significantly lower in the no-anchor condition than in the high-anchor condition, $t(198) = 2.54$, 95% CI for the mean difference = [13.21, 105.23], $p = .01$, $d = 0.36$. By contrast, there was no significant difference in dot estimates for the 35-dot or 97-dot arrays between the no-anchor condition and the high-anchor condition, both $t_s(198) < 0.91$, $p_s > .36$. Means and standard deviations are reported in Figure 2 (see also Table S1).

Experiment 3a: willingness to pay for Miami hotels

Method.

Participants and design. We requested 400 participants from MTurk, and 400 completed the experiment (47% female; age: $M = 37.30$ years, $SD = 11.79$). We randomly assigned each participant to a low- or high-externally-provided-anchor condition (between subjects) and to report the maximum amount they would be willing to pay for a two-star or four-star Miami hotel (between subjects).

Procedure. Participants imagined purchasing a hotel room for one night during an upcoming trip to Miami and were shown the name, a photograph, and a star rating for two hotels. In the low-anchor condition, participants first saw this information and the price of a room for one night in a one-star Miami Beach hotel (i.e., the Miami Beach International Hostel, priced at \$44). In the high-anchor condition, participants saw this information and the price of a room for one night in a five-star Miami Beach hotel (i.e., Four Seasons Hotel Miami, priced at \$610). On a separate page, participants were then shown this information for either a two-star or a four-star Miami Beach hotel (i.e., without prices) and reported the maximum amount they would be willing to pay for a room (U.S. dollars per night) in that hotel in an open-ended response box on that page.

Results. We tested our directional predictions, examining how much participants were willing to pay for the

two target hotels in a 2 (anchor: low, high) \times 2 (hotel: two star, four star) between-subjects ANOVA. It revealed significant main effects of anchor, $F(1, 396) = 175.56$, $p < .001$, $\eta_p^2 = .31$, and of hotel, $F(1, 396) = 137.36$, $p < .001$, $\eta_p^2 = .26$. More important, these main effects were qualified by a significant Anchor \times Hotel interaction, $F(1, 396) = 22.39$, $p < .001$, $\eta_p^2 = .05$. Participants were willing to pay more for the four-star hotel in the high-anchor condition than in the low-anchor condition, $t(198) = 10.63$, 95% CI for the mean difference = [164.78, 239.85], $p < .001$, $d = 1.50$. Participants were also willing to pay more for the two-star hotel in the high-anchor condition than in the low-anchor condition, $t(198) = 7.99$, 95% CI for the mean difference = [72.17, 119.50], $p < .001$, $d = 1.13$. The significant interaction revealed that this difference between anchor conditions was significantly greater for the four-star than the two-star hotel. Means and standard deviations are reported in Figure 2 (see also Table S1).

Experiment 3b: counts in dot arrays

Method.

Participants and design. We requested 400 participants from MTurk, and 400 completed the experiment (46% female; age: $M = 37.23$ years, $SD = 12.21$). We randomly assigned each participant to a low- or high-externally-provided-anchor condition (between subjects) and to evaluate a 35-dot or 273-dot array (between subjects).

Procedure. In the low-anchor condition, participants first saw a 10-dot array. In the high-anchor condition, participants first saw a 500-dot array. In both conditions, participants were told the number of dots depicted in that anchor array (i.e., 10 or 500) and that all of the dots in the experiment were of the same size. Next, in an open-ended response box, participants estimated the number of dots in either a 35-dot or a 273-dot array.

Results. We tested our directional predictions, examining dot estimates in a 2 (anchor: low, high) \times 2 (dot array: 35, 273) between-subjects ANOVA. It revealed significant main effects of anchor, $F(1, 396) = 177.52$, $p < .001$, $\eta_p^2 = .31$, and of array, $F(1, 396) = 618.90$, $p < .001$, $\eta_p^2 = .61$. More important, there was a significant Anchor \times Array interaction, $F(1, 396) = 143.30$, $p < .001$, $\eta_p^2 = .27$. The dot estimate for the 273-dot array was significantly lower in the low-anchor condition than in the high-anchor condition, $t(195) = 12.63$, 95% CI for the mean difference = [199.33, 273.09], $p < .001$, $d = 1.80$. Moreover, the dot estimate for the 35-dot array was significantly lower in the low-anchor condition than in the high-anchor condition, $t(201) = 4.08$, 95% CI for the mean difference = [6.52, 18.75], $p < .001$, $d = 0.57$. However, the interaction revealed that this difference between conditions was significantly greater for the 273-dot array than the 35-dot

array. Means and standard deviations are reported in Figure 2 (see also Table S1).

Discussion

Anchoring effects increased with stimulus magnitude across a variety of anchors, judgments, and targets—for both novel (i.e., dot arrays) and familiar (i.e., hotels, dogs, and french fries) stimuli.

Experiments 4a and 4b: Stimulus Magnitudes and the Replicability of Anchoring Effects

In Experiments 4a and 4b, we tested whether our framework explains instances in which anchoring effects were found to be weak or did not replicate. In an adaptation of the paradigm of Jung and colleagues (2016), we tested anchoring effects in a pay-what-you-want paradigm on prices for one donut and for 12 donuts (original and new quantity, respectively). In an adaptation of the paradigm of Maniadis and colleagues (2014), we tested anchoring effects on how much money participants were willing to accept to listen to an unpleasant tone for 60 s, 180 s, and 300 s (original and two new durations, respectively). We expected to find weak or no anchoring effects at the original low stimulus magnitudes but to find larger anchoring effects at the new higher stimulus magnitudes.

Experiment 4a: pay what you want for donuts

Method.

Participants and design. We requested 200 participants from MTurk, and 202 completed the experiment (40% female; age: $M = 36.39$ years, $SD = 10.03$). We randomly assigned each participant to one of two externally-provided-anchor conditions in a mixed design (low anchor or high anchor; between subjects). Each participant then indicated how much they would pay for one donut and a dozen donuts (within subjects).

Procedure. Participants saw the same images used to induce anchoring effects by Jung et al. (2016; i.e., Study 6a), which read “Dream Fluff Donuts! \$1 or Pay What You Want” (low anchor) and “Dream Fluff Donuts! \$3 or Pay What You Want” (high anchor). All participants then reported how much they would pay for a donut and a dozen donuts. Participants reported values for both donut purchases simultaneously on one survey page.

Results. Eleven participants were excluded from all analyses because their reported payment for one donut was higher than for a dozen donuts (preregistered

exclusion criterion). We first examined payments for the two donut purchases in a 2 (anchor: low, high; between subjects) \times 2 (quantity: one donut, one dozen donuts; within subjects) mixed ANOVA, which revealed a significant main effect of anchor, $F(1, 189) = 14.10$, $p < .001$, $\eta_p^2 = .07$, and a significant main effect of quantity, $F(1, 189) = 170.37$, $p < .001$, $\eta_p^2 = .47$. Simple comparisons found that participants would pay less for one donut in the low-anchor condition ($M = 1.61$, $SD = 1.94$) than in the high-anchor condition ($M = 2.53$, $SD = 2.32$), $t(189) = 2.94$, 95% CI for the mean difference = [0.30, 1.52], $p < .01$, $d = 0.43$, and would pay less for one dozen donuts in the low-anchor condition ($M = 8.67$, $SD = 8.75$) than in the high-anchor condition ($M = 14.35$, $SD = 12.60$), $t(189) = 3.60$, 95% CI for the mean difference = [2.56, 8.79], $p < .001$, $d = 0.52$. A significant Anchor \times Quantity interaction, $F(1, 189) = 10.843$, $p = .001$, $\eta_p^2 = .05$, however, showed that the anchoring effect was larger for a dozen donuts than for a single donut (see Fig. 3).

Experiment 4b: payment requested to listen to unpleasant tones

Method.

Participants and design. We requested 200 participants from MTurk, and 197 completed the experiment (42% female; age: $M = 37.80$ years, $SD = 10.87$). We randomly assigned each participant to one of two anchoring conditions in a mixed design (no anchor, low anchor; between subjects). Each participant reported their willingness to listen to an unpleasant tone for 60 s, 180 s, and 300 s (within subjects).

Procedure. In the no-anchor condition, participants saw no anchor. We first asked participants to put on their headphones (if they used them) and adjust their device volume (e.g., computer, smartphone) to a comfortable level. Participants then listened to a 30-s sample of an unpleasant tone, the same tone used by Maniadis et al. (2014). Next, they reported the minimum amount of money they would be willing to accept to listen to the same tone for 60 s, 180 s, and 300 s. Participants reported values for all three durations simultaneously on one survey page.

In the low-anchor condition, we asked participants, “Would you be willing to repeat the same experience for \$0.10? (Yes/No)” immediately after they listened to the tone sample for 30 s. They then reported their minimum amount of money for which they would be willing to listen to the same tone for 60 s, 180 s, and 300 s, just as controls. Finally, all participants responded to a manipulation check verifying that they listened to the sample tone—that is, “To which of the below was the sound most similar? (Police siren/Truck horn/Vacuum cleaner/High pitched beep).”

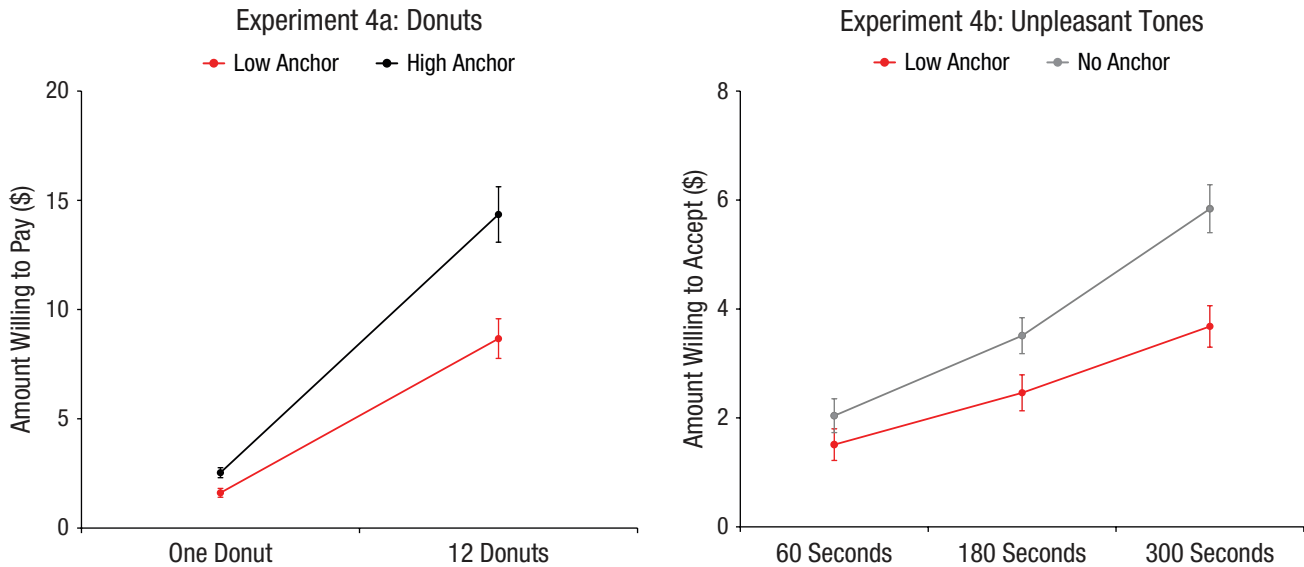


Fig. 3. Mean point estimate (*y*-axis) for each stimulus magnitude (*x*-axis) as a function of anchor condition in Experiments 4a and 4b. Participants indicated how much they wanted to pay for donuts (Experiment 4a) and the smallest amount of money they would be willing to accept to listen to an unpleasant tone (Experiment 4b). Error bars represent ± 1 SEM.

Results. Fifty participants were excluded from all analyses because they failed to correctly identify the tone as a “high pitched beep” (preregistered exclusion criterion). We first examined participants’ willingness to listen to tones over the three durations in a 2 (anchor: no, low; between subjects) \times 3 (duration: 60 s, 180 s, 300 s; within subjects) mixed ANOVA, which revealed a significant main effect of anchor, $F(1, 145) = 7.74, p < .01, \eta_p^2 = .05$, and a significant main effect of duration, $F(1, 145) = 137.83, p < .001, \eta_p^2 = .49$. More important, these main effects were qualified by a significant Anchor \times Duration interaction, $F(1, 145) = 10.55, p = .001, \eta_p^2 = .07$, suggesting that the anchoring effect increased with the duration of the unpleasant tone. Simple comparisons showed that participants were willing to accept less to listen to the tone for 300 s in the low-anchor condition ($M = \$3.68, SD = 3.27$) than in the no-anchor condition ($M = \$5.84, SD = 3.76$), $t(145) = 3.72, 95\% \text{ CI for the mean difference} = [1.01, 3.31], p < .001, d = 0.61$. Similarly, participants were willing to accept less to listen to the tone for 180 s in the low-anchor condition ($M = \$2.46, SD = 2.87$) than in the no-anchor condition ($M = \$3.51, SD = 2.80$), $t(145) = 2.26, 95\% \text{ CI for the mean difference} = [0.13, 1.98], p < .05, d = 0.37$. However, participants were not willing to accept less to listen to the tone for 60 s in the low-anchor condition ($M = \$2.04, SD = 2.62$) than in the no-anchor condition ($M = \$1.51, SD = 2.54$), $t(145) = 1.25, 95\% \text{ CI for the mean difference} = [-0.31, 1.37], p = .22, d = 0.21$ (see Fig. 3).

Discussion

Using stimuli from cases in which anchoring effects were found to be weak or did not replicate (Jung et al.,

2016; Maniadis et al., 2014), we found that anchoring effects again increased with stimulus magnitude. Anchoring effects were similarly weak or absent at the low stimulus magnitudes used originally, but anchoring effects were substantial at the new higher stimulus magnitudes. The results help identify the kinds of stimuli with which anchoring effects and similar phenomena (e.g., heuristics and biases) will replicate.

Experiment 5: Moderation by Anchor Relevance

Our theory predicts that anchoring effects increase with stimulus magnitude because the increased noise amplifies anchoring bias; noise does not itself induce anchoring effects. Because relevant anchors produce more anchoring bias than irrelevant anchors (Wilson et al., 1996), stimulus magnitude should increase anchoring effects more when anchors are relevant than irrelevant to stimulus estimates.

Method

Participants and design. We requested 600 participants from MTurk, and 600 completed the experiment (46% female; age: $M = 38.12$ years, $SD = 11.88$). We increased the sample size relative to previous experiments because of the number of levels in the between-subjects factor. In a mixed design, we randomly assigned each participant to a no-anchor condition, a relevant-internally-generated-low-anchor condition, or an irrelevant-internally-generated-low-anchor condition (between subjects). Each participant then reported the amount they

would be willing to pay for three Miami Beach hotels (within subjects).

Procedure. As in Experiment 1a, all participants imagined booking a hotel room for an upcoming trip to Miami.

In the hotel-anchor condition, participants first saw a picture, the name, and the rating for a one-star anchor Miami Beach hotel and reported how much they would be willing to pay (per night) for it. On a subsequent page, they saw a picture, the name, and the rating for each of three target Miami Beach hotels (two star, three star, and four star), as in Experiment 1a, and reported how much they would be willing to pay (per night) for one room in each hotel.

In the no-anchor condition, participants first reported how much they would be willing to pay (per night) for the three target hotels and then saw and reported how much they would be willing to pay (per night) for the one-star anchor hotel.

In the jeans-anchor condition, participants first saw the brand logo of Levi's and stated how much they would be willing to pay for one pair of Levi's jeans. They then saw the same information about each of the three target hotels and reported how much they would be willing to pay (per night) for one room in each hotel.

In all conditions, participants reported how much they would be willing to pay (U.S. dollars) for each stimulus in a unique open-ended response box.

Results

Anchor. An ANOVA examining how much participants would be willing to pay for the anchor itself across conditions revealed a marginal main effect of condition, $F(2, 597) = 2.79, p = .06$. Post hoc analyses revealed no significant difference in how much participants were willing to pay for the anchor between the no-anchor condition ($M = \$66.91, SD = 55.62$) and the hotel-anchor condition ($M = \$60.63, SD = 58.38; p = .30$). The amount they were willing to pay for the anchor in the jeans-anchor condition ($M = \$52.44, SD = 68.74$) was significantly lower than the amount they were willing to pay for the anchor in the no-anchor condition ($p = .02$). Most important, the amount they were willing to pay for the anchor in the jeans-anchor condition did not differ significantly from the amount they were willing to pay for the anchor in the hotel-anchor condition, $t(198) = 21.88, p < .18$.

Targets. We examined the moderating effect of anchor relevance on the amount participants were willing to pay for the target hotels in a 3 (anchor: no, hotel, jeans; between subjects) \times 3 (hotel: two star, three star, four star; within subjects) mixed ANCOVA with the amount participants were willing to pay for the anchor as a covariate. The analysis revealed a significant main effect

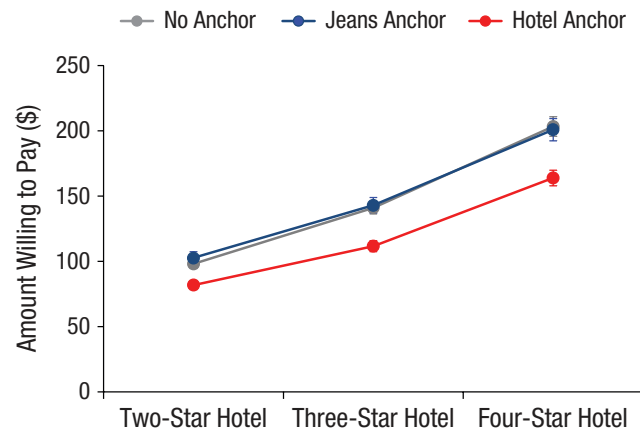


Fig. 4. Mean point estimate (y -axis) for each stimulus magnitude (x -axis) as a function of anchor condition in Experiment 5. Participants indicated how much they would be willing to pay per night for one room in three Miami Beach hotels. Error bars represent ± 1 SEM.

of anchor, $F(2, 596) = 15.76, p < .001, \eta_p^2 = .05$; a main effect of hotel, $F(1, 596) = 383.12, p < .001, \eta_p^2 = .39$; and a significant Anchor \times Hotel interaction, $F(2, 596) = 4.61, p = .01, \eta_p^2 = .02$. The amount participants were willing to pay for the anchor was a significant covariate, $F(1, 596) = 192.99, p < .001, \eta_p^2 = .25$. All results held when the amount participants were willing to pay for the anchor was not included as a covariate. Figure 4 shows the mean amount participants were willing to pay for the three focal hotels in each anchor condition.

We next decomposed this interaction in pairwise comparisons of conditions using separate 2×3 mixed ANCOVAs. Comparing the no-anchor and the hotel-anchor conditions revealed a significant 2 (anchor: no, hotel) \times 3 (hotel: two star, three star, four star) interaction, $F(1, 398) = 9.27, p = .002, \eta_p^2 = .02$. Simple comparisons showed that participants were willing to pay significantly more for the four-star hotel in the no-anchor condition ($M = \$203.29, SD = 105.09$) than in the hotel-anchor condition ($M = \$163.85, SD = 84.98$), $t(399) = 4.14, 95\%$ CI for the mean difference = $[20.71, 58.17], p < .001, d = 0.41$. They were also willing to pay more for the three-star and two-star hotels in the no-anchor condition than in the hotel-anchor condition—three star, no anchor: $M = \$141.00, SD = 66.57$; three star, hotel anchor: $M = \$111.67, SD = 60.45$; $t(399) = 4.62, 95\%$ CI for the mean difference = $[16.86, 41.81], p < .001, d = 0.46$; two star, no anchor: $M = \$98.04, SD = 47.61$; two star, hotel anchor: $M = \$81.83, SD = 53.27$; $t(399) = 3.21, 95\%$ CI for the mean difference = $[6.28, 26.14], p = .001, d = 0.32$. Moreover, a 2 (anchor: no, low) \times 2 (hotel: two star, four star) interaction revealed that this difference between conditions was significantly greater for the four-star than the two-star hotel, $F(1, 398) = 10.13, p = .002, \eta_p^2 = .03$.

Comparing the no-anchor and the jeans-anchor conditions in a 2 (anchor: no, jeans) \times 3 (hotel: two star,

three star, four star) ANCOVA found no interaction, $F(1, 393) = 0.14, p = .71, \eta_p^2 < .001$, suggesting that considering an irrelevant anchor did not increase anchoring effects with stimulus magnitude (two star, jeans anchor: $M = \$102.71, SD = 66.19$; three star, jeans anchor: $M = \$142.96, SD = 83.93$; four star, jeans anchor: $M = \$200.84, SD = 121.08$).

Finally, comparing the hotel-anchor and the jeans-anchor conditions in a 2 (anchor: hotel, jeans) \times 3 (hotel: two star, three star, four star) ANCOVA revealed a significant interaction, $F(1, 400) = 4.83, p = .029, \eta_p^2 = .01$. Simple comparisons revealed that participants were willing to pay significantly more in the jeans-anchor condition than in the hotel-anchor condition, $t(401) = 3.56, 95\% \text{ CI for the mean difference} = [16.54, 57.43], p < .001, d = 0.35$. Differences in the amount participants were willing to pay for the three-star and two-star hotels were also significant between the two anchor conditions, two-star hotel: $t(401) = 3.49, p = .001, 95\% \text{ CI for the mean difference} = [9.12, 32.63], d = 0.04$; three-star hotel: $t(401) = 4.30, p < .001, 95\% \text{ CI for the mean difference} = [17.00, 45.60], d = 0.19$. A 2 (anchor: hotel, jeans) \times 2 (hotel: two star, four star) interaction suggested that this difference between conditions was significantly greater for the four-star than the two-star hotel, $F(1, 400) = 5.09, p = .025, \eta_p^2 = .01$.

Discussion

Noise alone did not increase anchoring effects. Anchoring effects increased with stimulus magnitude only when the anchor was relevant to targets—when there was an anchoring bias to amplify.

Experiments 6a and 6b: Anchoring Effects, Anchoring Bias, and Noise

In Experiments 6a and 6b, we directly compared the effects of stimulus magnitude on anchoring effects, anchoring bias, and noise. We elicited range and point estimates for stimuli of small and large magnitude in low- and high-external-anchor conditions. We calculated range widths and anchoring effects, as before, but also calculated anchoring bias with a skew index (Epley & Gilovich, 2006). We predicted that stimulus magnitude would increase range widths and anchoring effects, even though anchoring bias would be similar for stimuli of both small and large magnitude.

Experiment 6a: willingness to pay for Miami hotels

Method.

Participants and design. We requested 400 participants from MTurk, and 407 completed the experiment

(49% female; age: $M = 38.51$ years, $SD = 12.20$). The design was adapted from Experiment 3a. We randomly assigned each participant to a low- or high-externally-provided-anchor condition (between subjects). Participants then made a range or point estimate involving how much they would be willing to pay for a standard room (per night) in either a two-star or a four-star Miami Beach hotel (all between subjects).

Procedure. All participants imagined vacationing in Miami in January 2023 after the pandemic ended. Participants randomly assigned to the low-anchor condition saw the one-star Miami Beach hotel and its \$44 rate, as in Experiment 3a. Participants randomly assigned to the high-anchor condition saw the five-star Miami Beach hotel and its \$610 rate, as in Experiment 3a. On the same page, participants also saw one target hotel: either the two-star or the four-star Miami Beach hotel from Experiment 3a (i.e., with no rate displayed). On a separate page, participants randomly assigned to a point-estimate condition then reported the maximum amount they were willing to pay for a room (U.S. dollars per night) in that target hotel in an open-ended response box. Participants randomly assigned to a range-estimate condition estimated the maximum and minimum possible average amount that the 100 participants who made a point estimate for the target hotel would be willing to pay. We had these participants estimate averages because the minimum and maximum possible individual estimates could range from zero to infinity. Range estimates were reported in two separate open-ended response boxes.

Results. All means and CIs are reported in Table 2.

Point estimates (anchoring effects). We examined point estimates in a 2 (anchor: low, high) \times 2 (hotel: two star, four star) between-subjects ANOVA. It revealed significant main effects of anchor, $F(1, 200) = 98.04, p < .001, \eta_p^2 = .33$, and hotel, $F(1, 200) = 104.72, p < .001, \eta_p^2 = .34$. More important, these main effects were qualified by a significant Anchor \times Hotel interaction, $F(1, 200) = 5.39, p = .02, \eta_p^2 = .03$. Participants were willing to pay more for the four-star hotel in the high-anchor condition than in the low-anchor condition, $t(100) = 6.86, 95\% \text{ CI for the mean difference} = [114.80, 208.16], p < .001, d = 1.38$. Participants were also willing to pay more for the two-star hotel in the high-anchor condition than in the low-anchor condition, $t(100) = 8.33, 95\% \text{ CI for the mean difference} = [76.29, 123.99], p < .001, d = 1.45$. The significant interaction revealed that this difference in point estimates between anchor conditions was significantly greater for the four-star than the two-star hotel.

Range estimates (noise). We converted range estimates to widths (i.e., maximum – minimum) and compared them

Table 2. Mean Range Estimates, Point Estimates, and Skew Index by Anchor and Target Magnitude in Experiments 6a and 6b

| Anchor and target | Range estimate | | Point estimate | Skew index |
|-------------------|---------------------------|---------------------------|---------------------------|----------------|
| | Minimum | Maximum | | |
| Experiment 6a | | | | |
| One-star hotel | | | | |
| Two-star hotel | \$57.21 [50.64, 63.78] | \$98.64 [85.65, 111.64] | \$72.69 [63.97, 81.41] | .37 [.17, .58] |
| Four-star hotel | \$118.15 [96.18, 140.13] | \$301.04 [236.74, 365.35] | \$177.21 [149.85, 204.58] | .32 [.17, .47] |
| Five-star hotel | | | | |
| Two-star hotel | \$117.43 [89.36, 145.51] | \$235.55 [191.90, 279.20] | \$172.83 [148.73, 196.93] | .53 [.33, .73] |
| Four-star hotel | \$227.07 [191.83, 262.31] | \$415.40 [368.62, 462.18] | \$338.69 [301.84, 375.54] | .41 [.21, .60] |
| Experiment 6b | | | | |
| 10 dots | | | | |
| 35 dots | 29.20 [26.38, 32.01] | 44.71 [40.33, 49.08] | 35.86 [33.48, 38.24] | .43 [.28, .58] |
| 273 dots | 109.42 [96.83, 122.01] | 205.54 [172.20, 238.88] | 145.02 [126.97, 163.07] | .37 [.19, .58] |
| 500 dots | | | | |
| 35 dots | 39.44 [31.36, 47.52] | 65.89 [54.46, 77.32] | 54.26 [44.29, 64.23] | .43 [.07, .80] |
| 273 dots | 303.46 [263.27, 343.65] | 467.30 [418.21, 516.40] | 425.30 [391.34, 459.26] | .25 [.05, .46] |

Note: Values in brackets are 95% confidence intervals.

in a 2 (anchor: low, high) \times 2 (hotel: two star, four star) between-subjects ANOVA. It revealed the predicted significant effect of hotel; ranges were wider for the four-star than two-star hotel, $F(1, 199) = 44.29, p < .001, \eta_p^2 = .18$. Exploratory analyses revealed that there was also a significant main effect of anchor, $F(1, 199) = 6.67, p = .01, \eta_p^2 = .03$ —ranges were wider in high-anchor than low-anchor conditions—and a significant Anchor \times Hotel interaction, $F(1, 199) = 5.02, p = .03, \eta_p^2 = .03$. Range estimates were still significantly wider for the four-star than the two-star hotel in the low-anchor condition (four star: $M = 182.89, SD = 180.69$; two star: $M = 41.43, SD = 34.76$), $t(100) = 5.74, p < .001, 95\% \text{ CI for the mean difference} = [92.54, 190.39], d = 1.09$, and in the high-anchor condition (four star: $M = 188.33, SD = 107.38$; two star: $M = 118.11, SD = 88.84$), $t(99) = 3.51, p = .001, 95\% \text{ CI for the mean difference} = [30.50, 109.94], d = 0.71$. We interpret these wider ranges with higher than lower anchors as further evidence of the influence of scalar variability. In other words, the larger anchor may have made the value of the hotels appear greater and thus increased the noise in their estimation, but this interpretation is admittedly speculative.

Skew index (anchoring bias). We then calculated a skew index (Epley & Gilovich, 2006) to quantify anchoring bias across point-estimate conditions. We divided (a) the difference between each participant's point estimate and the range end point nearest to the anchor by (b) the total range width of plausible values: (point estimate – the maximum or minimum plausible value)/(the maximum plausible value – the minimum plausible value). The range-estimate values used for each participant were specific to their treatment (e.g., low anchor, two-star hotel).

We then multiplied the skewness index in the high-anchor conditions by -1 so adjustment could be directly compared with the low-anchor conditions. Generally, a lower skew index suggests greater anchoring bias. Perfectly centered estimates received a score of .50. Estimates closer to the anchor, cases of insufficient correction from the anchor, scored less than .50. Estimates beyond the midpoint, cases of overcorrection from the anchor, scored higher than .50.

We examined skew indices in a 2 (anchor: low, high) \times 2 (hotel: two star, four star) between-subjects ANOVA. It revealed no significant main effect of anchor, $F(1, 200) = 1.55, p = .22, \eta_p^2 < .01$; no main effect of hotel, $F(1, 200) = 0.83, p = .36, \eta_p^2 < .01$; and no Anchor \times Hotel interaction, $F(1, 200) = 0.15, p = .70, \eta_p^2 < .01$.

Experiment 6b: counts in dot arrays

Method.

Participants and design. We requested 400 participants from MTurk, and 400 completed the experiment (46% female; age: $M = 39.59$ years, $SD = 12.42$). The paradigm was the same as in Experiment 6a, and the dot arrays used in Experiment 3b were used as anchors and targets. We randomly assigned each participant to a low- or high-externally-provided-anchor condition (between subjects). Participants then made a range or point estimate for a 35-dot or 273-dot array (all between subjects).

Procedure. As in Experiment 3b, participants randomly assigned to the low-anchor condition saw a 10-dot array. Participants randomly assigned to the high-anchor condition saw a 500-dot array. Participants were told the number of dots in that array (i.e., 10 or 500) and that all

dots in the experiment were the same size. Next, each participant was randomly assigned to see a 35-dot or a 273-dot array (the number of dots was not labeled). Participants randomly assigned to a point-estimate condition estimated the number of dots in the array in an open-ended response box. Participants randomly assigned to a range-estimate condition estimated the maximum and minimum plausible number of dots in the array. Range estimates were reported in two separate open-ended response boxes.

Results. All means and CIs are reported in Table 2.

Point estimates (anchoring effects). We examined point estimates in a 2 (anchor: low, high) \times 2 (dot array: 35, 273) between-subjects ANOVA. It revealed a significant main effect of anchor, $F(1, 196) = 241.14, p < .001, \eta_p^2 = .55$; a significant main effect of dot array, $F(1, 196) = 623.284, p < .001, \eta_p^2 = .76$; and the predicted significant Anchor \times Dot Array interaction, $F(1, 196) = 185.37, p < .001, \eta_p^2 = .49$. Participants estimated a significantly lower count for the 273-dot array in the low-anchor condition than in the high-anchor condition, $t(99) = 15.19, 95\% \text{ CI for the mean difference} = [243.67, 316.89], p < .001, d = 2.98$. Participants also estimated a significantly lower count for the 35-dot array in the low-anchor condition than in the high-anchor condition, $t(97) = 4.14, 95\% \text{ CI for the mean difference} = [9.57, 27.24], p < .001, d = 0.78$. The significant interaction revealed that the anchoring effect on point estimates was larger for the 273-dot array than the 35-dot array.

Range estimates (noise). We converted range estimates to widths (i.e., maximum – minimum) and compared them in a 2 (anchor: low, high) \times 2 (dot array: 35, 273) between-subjects ANOVA. Most important, it revealed the predicted significant effect of dot array, $F(1, 196) = 108.82, p < .001, \eta_p^2 = .37$. Exploratory analyses revealed that there was also a significant main effect of anchor, $F(1, 196) = 14.17, p < .001, \eta_p^2 = .07$ —ranges were wider in the high-anchor than low-anchor conditions—and a significant Anchor \times Dot Array interaction, $F(1, 196) = 7.39, p < .01, \eta_p^2 = .04$. Range estimates were still significantly wider for the 273-dot array than the 35-dot array in the low-anchor condition (273-dot array: $M = 96.12, SD = 84.25$; 35-dot array: $M = 15.51, SD = 14.59$), $t(89) = 5.84, p < .001, 95\% \text{ CI for the mean difference} = [53.20, 108.02], d = 1.29$, and in the high-anchor condition (273-dot array: $M = 163.85, SD = 115.73$; 35-dot array: $M = 26.44, SD = 31.05$), $t(107) = 9.00, p < .001, 95\% \text{ CI for the mean difference} = [107.15, 167.65], d = 1.62$. As for Experiment 6a, we interpret the wider ranges with higher than lower anchors as further evidence of the influence of scalar variability. The larger anchor may have made the arrays appear larger in number and thus increased the noise

in participants' estimates, but again this interpretation is admittedly speculative.

Skew index (anchoring bias). Skew index was calculated using the same method as Experiment 6a. We examined skew index in a 2 (anchor: low, high) \times 2 (dot array: 35, 273) between-subjects ANOVA, which revealed no significant main effect of anchor, $F(1, 196) = 0.24, p = .62, \eta_p^2 < .01$; no main effect of dot array, $F(1, 196) = 1.11, p = .29, \eta_p^2 < .01$; and no Anchor \times Dot Array interaction, $F(1, 196) = 0.30, p = .59, \eta_p^2 < .01$.

Discussion

Anchoring effects on point estimates and ranges of plausible values increased with stimulus magnitude. Anchoring bias as measured by the skew index, however, was not statistically different for the hotels or dot arrays with small and large magnitudes ($F_s \leq 1.11, p_s \geq .292$). Anchoring effects appear to have increased with stimulus magnitude because of increased noise in stimulus representations and not because of increased anchoring bias.

General Discussion

Anchoring effects increased with stimulus magnitude. This appears to have been because of an increase in judgmental noise. Ranges of plausible values for point estimates increased with stimulus magnitudes, but anchoring bias did not. As scalar variability would predict, regressing standard deviations on means of all target estimates also revealed a positive linear relationship between the noise and magnitude of point estimates, $\beta_1 = 0.44, SE = 0.02; t(60) = 19.96, p < .001; F(1, 60) = 398.32, p < .001, R^2 = .87$ (see Fig. 5).

Alternative explanations, such as a floor effect of scales or an inability to differentiate low anchors from low-magnitude stimuli, are not supported by the data. The lower bound of plausible ranges for all low-magnitude stimuli (i.e., average minimum) was significantly greater than zero (all $t_s \geq 3.37$, all $p_s \leq .001$) and significantly greater than all low anchors (all $t_s \geq 2.53$, all $p_s \leq .015$; full statistics are reported in Section S5 in the Supplemental Material). Anchoring bias induced by low and high anchors did not differ for the two-star hotel in Experiment 6a or the small dot array in Experiment 6b ($t_s < 1, p_s \geq .283$). An ancillary experiment found that the effect of stimulus magnitude on anchoring effects held when stimulus values were negative integers (see the Supplemental Material), and comparisons between the no-anchor and high-anchor conditions in Experiments 1 and 2b (see also Section S3 in the Supplemental Material), where censoring effects

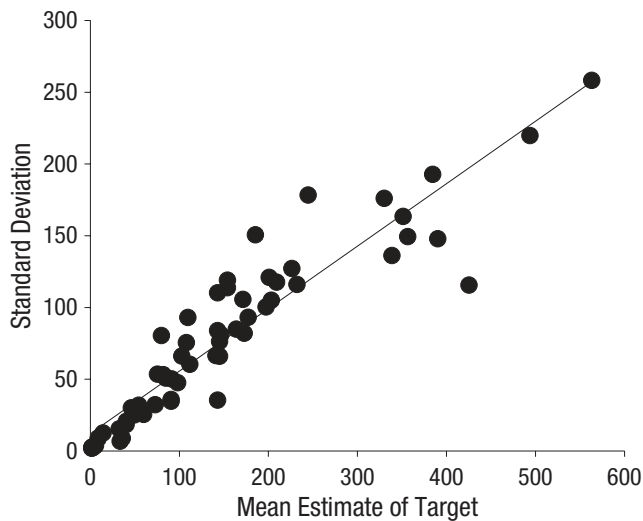


Fig. 5. Scatterplot showing the standard deviation of point estimates by the mean of point estimates for all targets in all experiments. The solid line shows the best-fitting regression.

should not apply, also found an increase in anchoring effects with stimulus magnitude. Of course, the underlying driver of our effect is judgmental noise, for which stimulus magnitude serves as a proxy. Important boundaries of our predictions should be found in contexts in which other factors lead stimulus magnitude and judgmental noise to be uncorrelated or negatively correlated. There, the best determinant of noise rather than stimulus magnitude should best predict the size of anchoring effects. Anchoring effects for prices should be greater for lower denominations of currencies that are less familiar to buyers and sellers (e.g., Bitcoin or pounds) than larger denominations of more familiar currencies (e.g., dollars or euros), for instance, and greater for health-care professionals estimating case numbers for rare and unusual new viruses (e.g., COVID-19) than prevalent viruses with which they are more familiar (e.g., seasonal influenza).

Our findings elucidate the roles of noise and bias in anchoring effects. Noise modulates the size of anchoring effects by modulating anchoring bias. Our findings and framework contribute to the anchoring literature by reconciling questions regarding the replicability and prevalence of anchoring effects (Jung et al., 2016; Maniadi et al., 2014). The stimuli used in that research were so low in magnitude that they induced insufficient noise to observe sizeable anchor effects. More important, our findings suggest a reexamination of how anchoring effects are moderated by factors such as cognitive load, intoxication, subjective confidence, knowledge, and incentives (Epley & Gilovich, 2006; Jacowitz & Kahneman, 1995; Mussweiler & Strack, 2000; Simmons et al., 2010; Smith et al., 2013). It is often assumed that they

modulate anchoring effects by influencing anchoring bias, but these factors could also modulate anchoring effects by influencing judgmental noise. Cognitive load or intoxication, for instance, may increase anchoring effects in point estimates by reducing adjustment from an anchor (i.e., anchoring bias) or by widening the range of values perceived to be plausible (i.e., noise). More broadly, our framework shows how noise can modulate effects of heuristics and biases on judgments under uncertainty, and it provides a paradigm for testing the role of noise in these phenomena.

Transparency

Action Editor: Leah Somerville

Editor: Patricia J. Bauer

Author Contributions

Both authors contributed equally to the study design. C.-Y.

Lee collected and analyzed the data under the supervision of C. K. Morewedge. Both authors contributed equally to the preparation of the manuscript and approved the final manuscript for submission.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Open Practices

All data and materials have been made publicly available via OSF and can be accessed at <https://osf.io/9xun6/>. The design and analysis plan for each experiment was preregistered on AsPredicted.org (Experiment 1a: <https://aspredicted.org/g3sf5.pdf>; Experiment 1b: <https://aspredicted.org/xe5zx.pdf>; Experiment 2a: <https://aspredicted.org/xs4as.pdf>; Experiment 2b: <https://aspredicted.org/3x4as.pdf>; Experiment 3a: <https://aspredicted.org/cp5aq.pdf>; Experiment 3b: <https://aspredicted.org/e6c97.pdf>; Experiment 4a: <https://aspredicted.org/3ye6y.pdf>; Experiment 4b: <https://aspredicted.org/hv3cd.pdf>; Experiment 5: <https://aspredicted.org/cc6zs.pdf>; Experiment 6a: <https://aspredicted.org/9ri7p.pdf>; Experiment 6b: <https://aspredicted.org/hb2ck.pdf>).

This article has received the badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.



ORCID iDs

Chang-Yuan Lee  <https://orcid.org/0000-0002-8600-5202>

Carey K. Morewedge  <https://orcid.org/0000-0001-7502-9279>

Acknowledgments

We thank Gretchen Chapman, Fiery Cushman, Nicholas Epley, Daniel Kahneman, Thomas Mussweiler, and Lawrence Williams for helpful feedback.

Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/09567976211024254>

Note

1. Note that in all mixed and within-subjects analyses in this article, we report the lower-bound (i.e., the most conservative) estimates.

References

- Chapman, G. B., & Johnson, E. J. (1994). The limits of anchoring. *Journal of Behavioral Decision Making*, 7(4), 223–242.
- Englich, B., Mussweiler, T., & Strack, F. (2006). Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making. *Personality and Social Psychology Bulletin*, 32(2), 188–200.
- Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science*, 17(4), 311–318.
- Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, 8(7), 307–314.
- Jacowitz, K. E., & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin*, 21(11), 1161–1166.
- Jung, M. H., Perfecto, H., & Nelson, L. D. (2016). Anchoring in payment: Evaluating a judgmental heuristic in field experimental settings. *Journal of Marketing Research*, 53(3), 354–368.
- Kahneman, D., Rosenfield, A. M., Gandhi, L., & Blaser, T. (2016). Noise: How to overcome the high, hidden cost of inconsistent decision making. *Harvard Business Review*, 94(10), 38–46.
- Maniadis, Z., Tufano, F., & List, J. A. (2014). One swallow doesn't make a summer: New evidence on anchoring effects. *American Economic Review*, 104(1), 277–290.
- Mussweiler, T., & Strack, F. (2000). Numeric judgments under uncertainty: The role of knowledge in anchoring. *Journal of Experimental Social Psychology*, 36(5), 495–518.
- Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: A problem of significance. *Nature Neuroscience*, 14(9), 1105–1107.
- Northcraft, G. B., & Neale, M. A. (1987). Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes*, 39(1), 84–97.
- Quattrone, G. A., Lawrence, C. P., Finkel, S. E., & Andrus, D. C. (1984). *Explorations in anchoring: The effects of prior range, anchor extremity, and suggestive hints* [Unpublished manuscript]. Stanford University.
- Simmons, J. P., LeBoeuf, R. A., & Nelson, L. D. (2010). The effect of accuracy motivation on anchoring and adjustment: Do people adjust from provided anchors? *Journal of Personality and Social Psychology*, 99(6), 917–932.
- Smith, A. R., Windschitl, P. D., & Bruchmann, K. (2013). Knowledge matters: Anchoring effects are moderated by knowledge level. *European Journal of Social Psychology*, 43(1), 97–108.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131.
- Wilson, T. D., Houston, C. E., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: Basic anchoring and its antecedents. *Journal of Experimental Psychology: General*, 125(4), 387–402.